Machine learning-based jet and event classification at the Electron-Ion Collider and applications to hadron structure and spin physics

Kyle Lee,[1,2,*] James Mulligan,[1,3,†] Mateusz Płoskoń,[1,‡] Felix Ringer,[4,5,6,7,§] and Feng Yuan[1,¶]

[1] *Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[2] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[3] *Physics Department, University of California, Berkeley, CA 94720, USA*
[4] *C.N. Yang Institute for Theoretical Physics, Stony Brook University, Stony Brook, NY 11794, USA*
[5] *Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA*
[6] *Department of Physics, Old Dominion University, Norfolk, Virginia 23529, USA*
[7] *Thomas Jefferson National Accelerator Facility, 12000 Jefferson Avenue, Newport News, Virginia 23606, USA*

We explore machine learning-based jet and event identification at the future Electron-Ion Collider (EIC). We study the effectiveness of machine learning-based classifiers at the relatively low EIC energies, focusing on (i) identifying the quark flavor of the jet and (ii) identifying the hard-scattering process. We propose applications of our machine learning-based jet identification in the key research areas at the future EIC and current RHIC program, including the extraction of (transverse momentum dependent) parton distribution functions, improving experimental access to transverse spin asymmetries, and quantifying the modification of hadrons and jets in the cold nuclear matter environment in electron-nucleus collisions. We establish first benchmarks and contrast the performance of flavor tagging at the EIC with that at the Large Hadron Collider. We perform studies relevant for the detector design including particle identification, charge information, and minimum transverse momentum requirements. Additionally, we study the impact of using full event information instead of using only information associated with the identified jet.

CONTENTS

* kylel@mit.edu
† james.mulligan@berkeley.edu
‡ mploskon@lbl.gov
§ fmringer@jlab.org
¶ fyuan@lbl.gov

I. INTRODUCTION

The future Electron-Ion Collider (EIC) will map out in detail the structure of nucleons and nuclei and allow for novel studies of hadronization, jets and cold nuclear matter [1]. The EIC center-of-mass (CM) energy of up to $\sqrt{s} = 141$ GeV will allow for detailed QCD studies using jets, which are highly energetic collimated sprays of particles observed in the detector. The measured energy and the direction of a jet represent good proxies of the corresponding quantities at the level of quarks and gluons that initiate the observed jet. This close correspondence between partons and jets was explored in Refs. [2, 3] for the expected EIC kinematics. In this work, we explore the use of machine learning to classify different jets as well as entire events using A.I. and machine learning. In addition, we outline specific applications that we envision at the future EIC. We expect that the tools and applications discussed in this work are also relevant for the ongoing experimental program at the Relativisitic Heavy Ion Collider (RHIC).

The rapid progress in A.I. and machine learning over the last decade has led to various applications in nuclear and high energy physics. See Ref. [4, 5] and references therein for recent developments in the areas of classification, generative modeling [6–8], regression and inference. In the context of the EIC, machine learning techniques have been proposed for example to determine kinematic

variables in Deep Inelastic Scattering (DIS) [9, 10] and to extract quantum correlation functions such as PDFs, TMDs, GPDs, and fragmentation functions [11**?**].

In high-energy collider physics, the classification of quark vs. gluon jets or QCD vs. W/Z jets has been studied with increasing sophistication over the past years [12]. An important goal of these studies is to increase the sensitivity to potential signals of physics beyond the Standard Model. Examples of machine learning architectures to classify jets include Convolutional Neural Networks (CNNs) [13, 14], deep sets [15], and transformer models [16]. Compared to tagging algorithms based on traditional observables, algorithms based on supervised machine learning algorithms have both benefits and drawbacks. The principal benefit is the ability of the machine learning algorithm to take advantage of the full information at hand and thereby significantly outperform algorithms based on traditional observables. The drawbacks in doing so are that the results can be difficult to interpret and connect to first principles since it is unclear what the machine is learning, and that the (simulated or experimental) data used in the training process may contain biases. At the LHC, machine learning based jet taggers have succeeded in significantly outperforming traditional jet taggers. At the same time, numerous efforts have been taken to address the limitations of machine learning based algorithms, both to improve the interpretability of machine learned results and bring under control the biases of the training data set. For example, in Refs. [17–20] complete bases of jet substructure observables were introduced which span the phase space of emissions inside jets. This large set of observables can rival the performance of successful machine learning algorithms, which has allowed for insights into how machines learn and an improved understanding of their good performance. In this way, machine learning based algorithms have driven progress both in the performance of the classification tasks but also in pushing traditional approaches forward. A key question at the EIC is understanding how large an improvement in performance machine learning based tagging algorithms can provide, and to identify applications of machine learning based algorithms that can drive forward the physics goals of the EIC, as it has at the LHC.

In this work, we will address the following three topics. First, we explore the application of machine learning-based classifiers at the comparatively low EIC energies. Typically, jet classification studies are carried out at LHC energies where the jet's transverse momentum is $\mathcal{O}(100$ GeV). Instead, at the EIC, jets will be produced predominantly with 10-30 GeV. Moreover, an important aspect of machine learning for jets is the sparsity of the data compared to typical tasks in computer vision. Due to the relatively low number of hadrons that make up EIC jets, we expect an increased level of sparsity compared to LHC jets. We will explore if the reduced energy and increased sparsity of jets at the EIC affect the performance of machine learning based classifiers.

Second, we explore both quark flavor jet tagging and quark vs. gluon jet tagging. As examples, we consider the binary classification tasks: u vs. d quark [21], ud vs. s and uds vs. c jet tagging and we compare to the jet charge [22, 23] as a reference. We expect that the flavor tagging of jets will be an important component to constrain collinear and transverse momentum dependent parton distribution functions (PDFs), which we discuss more below.

Third, we will investigate how the performance can be improved by not only making use of the particles inside the jet but also out-of-jet particles to classify the hard-scattering event. Out-of-jet radiation patterns include subleading jets or soft particles in the event. By making use of this additional information, we extend the jet classification task to event classification. For typical applications that we foresee at the future EIC, the classification of the hard-scattering event can improve measurements that are not necessarily limited to in-jet dynamics. Event-wide classification algorithms can help to improve the measurement of spin asymmetries and studies of cold nuclear matter. The machine learned event-wide information can also be mapped to traditional observables like N-jettiness ratios [24] or the jet pull [25, 26].

We foresee several specific applications of machine learning based jet and event classification to some of the major physics goals of the EIC:

(i) Strengthening constraints on (transverse momentum dependent) PDFs. We expect that the flavor tagging of jets will be an important component to constrain collinear and transverse momentum dependent parton distribution functions (PDFs), which we discuss more below. For example, in Ref. [27] charm-tagged jets were proposed to increase the sensitivity to the (collinear) strange quark PDF in charged current events. In Ref. [28], the authors considered jet substructure observables to constrain the gluon PDF at the LHC. See also Refs. [29–33] for further theoretical work on the flavor of jets. We generally find that machine learning-based classifiers outperform traditional observables like the jet charge and therefore we expect that machine learning can significantly enhance the science potential of the EIC. Directly including machine-learned event-by-event classifiers in a global analysis of quantum correlation functions like PDFs may require additional investigations, which are beyond the scope of this work. Alternatively, the machine learned classifier can be considered as an upper bound on the information content [18] and subsequently a suitable high-level observable that is calculable in perturbative QCD may be identified using for example lasso regression [16, 34] or other symbolic regression techniques [35].

(ii) Enhancing transverse single spin asymmetries. Transverse Single Spin Asymmetries (TSSAs) constitute some of the hallmark measurements at RHIC and the future EIC and they provide constraints on the spin structure of the proton. TSSAs are defined as the difference of cross sections where the incoming protons have different

transverse spin orientations

$$A_{UT} = \frac{\mathrm{d}\sigma^{\uparrow} - \mathrm{d}\sigma^{\downarrow}}{\mathrm{d}\sigma^{\uparrow} + \mathrm{d}\sigma^{\downarrow}} \,. \tag{1}$$

However, due to experimental constraints, it has generally been challenging to measure non-zero TSSAs. See Ref. [**?**] for recent measurements by the STAR Collaboration in $p^{\uparrow} + p \to$ di-jet+X correlation events where the jet charge was used as an additional measurement to increase the size of the asymmetry. See also Refs. [36–41] for further experimental results. Here we propose that an enhancement of TSSAs

$$\max_{\theta} A_{UT}(\theta) \,, \tag{2}$$

can be achieved by including an additional machine-learned measurement, which is here given by the parameters θ. This can be achieved by formulating this regression task as a classification problem of jets (or events) that are obtained in scattering processes with differently polarized protons ($\uparrow\downarrow$) in the initial state. By applying a classifier trained on distinguishing jets in events with different initial spin orientations as an additional measurement, similar to the jet charge, larger spin asymmetries may be obtained, which can provide better constraints on the corresponding quantum correlation functions in global analyses.

(iii) Elucidating cold nuclear matter effects. One of the goals of the EIC is to achieve an understanding of the transport coefficients of nuclear matter such as by comparing observables in eA collisions to those in ep collisions, similar to the jet quenching program comparing AA and pp collisions at RHIC and the LHC. The entire basis for extracting such properties of nuclear matter is the difference between eA and ep observables. By training machine learning methods to distinguish these two classes of events, one can use interpretable machine learning methods to gain insight into the type of information responsible for these differences, and thereby make connections to calculable observables in perturbative QCD [34].

The remainder of this work is organized as follows. In section III, we discuss the event generation and present the different machine learning algorithms used in this work. In section IV we present results for jet flavor classification at the EIC. In section V, we extend the jet classification to full events and compare the performance of machine learned results. In section VI, we draw conclusions and provide an outlook. (Need to list "theory section" here.)

II. MACHINE LEARNING APPLICATIONS TO HADRON STRUCTURE AND SPIN PHYSICS

A. Applications of jet tagging

Due to the large cancellations between different PDFs and fragmentation functions, jet flavor classification may help to significantly improve the measurement of spin-dependent observables. We note that the techniques discussed here do not only apply to TSSAs but they can be applied are more generally:

- Quark flavor and quark vs. gluon jet identification can help to improve the sensitivity to the longitudinally polarized gluon distribution Δg. See Refs. [42–46] for recent discussions and experimental results.
- Quark vs. gluon jet classification may help to improve measurements of the gluon Sivers function at RHIC and the future EIC [47].
- The techniques discussed here may also improve searches of physics beyond the Standard Model at the EIC [48–50]. For example, in Ref. [51] jet charge weighted TSSAs were proposed in this context.
- Exclusive, diffractive processes with jets, GPDs, Wigner functions [52–54] ... Still need to extend this and can also write a whole paragraph about it.

We leave a more detailed exploration of machine learning applications for these topics for future work [55].

B. Enhancing spin asymmetries

Transverse Single Spin Asymmetries (TSSA) where an incoming proton is transversely polarized

$$A_{UT} = \frac{\mathrm{d}\sigma^{\uparrow} - \mathrm{d}\sigma^{\downarrow}}{\mathrm{d}\sigma^{\uparrow} + \mathrm{d}\sigma^{\downarrow}} \,, \tag{3}$$

are an important component of the spin physics program at RHIC and the future EIC. However, experimental measurements are challenging and due to the relatively large uncertainties the asymmetries are often small or consistent with zero. Examples at RHIC include asymmetries in di-jet correlations $p^{\uparrow} + p \to \text{dijets} + X$. See Refs. [37**?**] for measurements from the STAR Collaboration and Refs. [56–58] for recent theoretical results. Other examples include single-inclusive measurements of pions $p^{\uparrow} + p \to \pi + X$. Similar measurements have been performed using jets, open heavy-flavor mesons and photons. See Refs. [38–41] for recent from the STAR and PHENIX Collaborations.

The reason that these asymmetries are small is the cancellation of contributions with opposite sign from different partons. Within QCD factorization, this can be traced back to PDFs or fragmentation functions of partons with different flavor. We will discuss two examples in more detail below where for example different quark PDFs contribute to the asymmetry with opposite sign and similar magnitude. Therefore, in Ref. [**?**], the STAR collaboration used the jet charge to tag jets with a specific quark flavor and a non-vanishing asymmetry was obtained. As demonstrated also in previous sections, machine learning-based jet or event classification outperforms traditional

observables. Therefore, we propose to enhance the measurement of TSSAs using an additional machine learned classifier or observable

$$\max_{\theta} A_{UT}(\theta)\,. \tag{4}$$

Here the additional machine learned classifier is given in terms of the set of parameters θ. The regression problem in Eq. (4), can be formulated as a classification task where the machine learning techniques discussed in previous sections can be directly applied. By training a classifier that distinguishes jets produced in events where the incoming proton has opposite transverse polarization ($\uparrow\downarrow$), we can find a classifier that maximizes the TSSA A_{UT}. This approach is similar to the discrimination task between jets in proton-proton and heavy-ion collisions where the trained classifier can be used to maximize the nuclear modification factor

$$\max_{\theta} R_{AA}(\theta) = \max_{\theta} \frac{\mathrm{d}\sigma_{AA}}{\mathrm{d}\sigma_{pp}}(\theta)\,. \tag{5}$$

See Refs. [] for more details. The identification of a machine learned-classifier can be performed directly on data before unfolding [34] or using corrected full events [59]. Subsequently, an observable can be identified that is calculable in perturbative QCD and that approximates the performance of the machine learned-classifier. For example, this can be achieved using complete sets of observables such as the N-subjettiness basis or EFPs discussed above. Using this observable, an unfolding procedure can be applied and the data can eventually be included in a global analyses of quantum correlation functions. We note that the target or loss function of the regression problem in Eq. (4) is at the hadron or observable level. However, within QCD factorization we can establish a direct connection to parton-level PDFs or fragmentation functions, which we will discuss in the following for two specific examples.

As discussed above, machine learning based classifiers that can distinguish jets or events that originate from different quark flavors (and gluons) can assist the spin physics program at RHIC and the future EIC. In particular, measurements of Sivers [60] and Collins [61] asymmetries are often close to zero due to large cancellations between different PDFs and fragmentation functions. These approximate cancellations can be understood from momentum sum rules. In the following, we will consider the Schäfer-Teryaev sum rule [62, 63] and the Burkardt [64–66] sum rule that are satisfied by the Collins and Sivers functions, respectively. Both of these sum rules state that average transverse momentum should sum to zero when summed over either the outgoing hadron flavors (Collins) or incoming quark flavors (Sivers). We note that the derivation of these sum rules involves bare quark and gluon operators and it is therefore unknown how much the sum rules are violated due to renormalization. Nevertheless, they provide an intuitive understanding of the large cancellations between different quark flavors to first order.

First, we consider the Sivers function $f_{1T}^{\perp a}(x, \vec{k}_T^2)$, which describes the longitudinal x and transverse momentum k_T anisotropy of partons inside a transversely polarized proton. Here the superscript $a = q, \bar{q}, g$ denotes the parton inside the proton. Including appropriate prefactors and formally integrating over the transverse momentum dependence, we find

$$f_{1T}^{\perp(1)a}(x) = \int \mathrm{d}^2\vec{k}_T \frac{\vec{k}_T^2}{2M^2} f_{1T}^{\perp a}(x, \vec{k}_T^2)\,. \tag{6}$$

The Burkardt sum rule for the Sivers function states that the following integral vanishes []

$$\sum_{a=q,\bar{q},g} \int_0^1 \mathrm{d}x f_{1T}^{\perp(1)a}(x) = 0\,. \tag{7}$$

Under the assumption that the valence quarks distributions dominate, the Burkardt sum rule leads to u and d-quark Sivers function that have opposite sign and similar magnitude. This expected behavior of the u and d quark Sivers functions has been confirmed by several global analyses [**?**].

Second, we consider the Collins fragmentation function $H_{1,h/q}^{\perp}(z, \vec{P}_\perp^2)$ as an example of spin-dependent dynamics in the final state where large flavor cancellations are expected. It describes the longitudinal z and transverse momentum $P_\perp$ distribution of a final-state hadron that originates from a transversely polarized parton. After integrating out the transverse momentum dependence, we find

$$H_{1,h/q}^{\perp(3)}(z) = \int \mathrm{d}^2\vec{P}_\perp \frac{\vec{P}_\perp^2}{M_h} H_{1,h/q}^{\perp}(z, \vec{P}_\perp^2)\,. \tag{8}$$

The Schäfer-Teryaev sum rule for the Collins function states that the integral over the longitudinal momentum fraction vanishes after we sum over all hadron species []

$$\sum_h \int_0^1 \mathrm{d}z\, H_{1,h/q}^{\perp(3)}(z) = 0\,. \tag{9}$$

For simplicity, we now assume that isospin symmetry holds and we assume that the light partons-to-pion fragmentation process dominates. In this case, only two fragmentation channels remain. The *favored* fragmentation functions are pion fragmentation functions for a valence parton u or d, i.e. $H_{1,\pi^+/u}^{\perp}$ or $H_{1,\pi^-/d}^{\perp}$, respectively, and the *unfavored* fragmentation functions are pion fragmentation functions of u or d that are not a valence partons, i.e. $H_{1,\pi^-/u}^{\perp}$ or $H_{1,\pi^+/d}^{\perp}$). According to the Schäfer-Teryaev sum rule, these two channels then need to have opposite sign. Therefore, experimental measurements are traditionally carried out with identified hadrons to achieve a flavor separation and avoid the approximate cancellation of different fragmentation functions. Achieving a flavor separation using the jet charge was investigated in Refs. [67, 68] and, analogously, we expect that machine

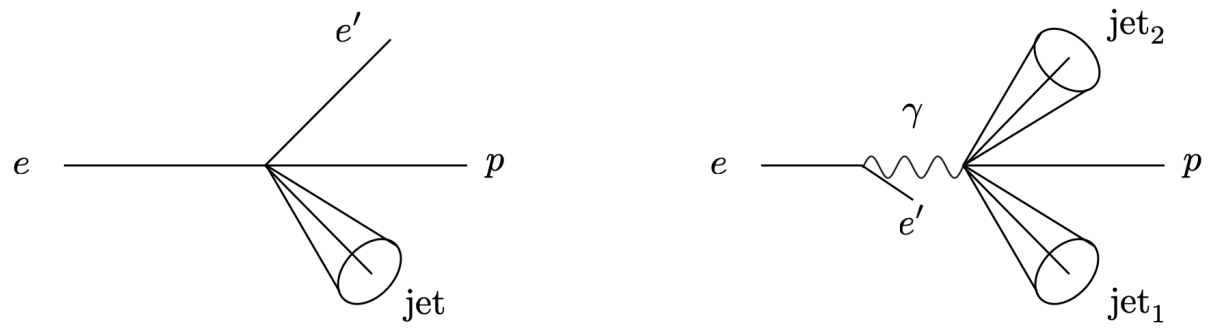

FIG. 1. Illustration of the jet production processes considered in this work. Left: High-Q^2 electron-proton scattering. At leading order, the final state consists of the scattered electron and a single jet originating from different quark flavors. Right: Low-Q^2 photoproduction, where we include both the direct and the resolved contribution. At leading order, the final state consists of the scattered electron in the forward direction / close to the beam axis and a di-jet pair, which can be initiated by both quarks and gluons. In both cases, the transverse momentum of the jets is measured relative to the beam axis in the laboratory frame.

learning-based observables can lead to new opportunities. Moreover, multi-differential measurements of identified hadrons and machine learning-based observables may also provide new insights, which we plan to investigate in the future.

III. SIMULATION AND TRAINING SETUP

To perform our studies, we generate simulated events using the Monte Carlo event generator PYTHIA6 [], which serves as the training data for the (supervised) machine learning based classification algorithms. In the following, we describe the simulated event sample and the machine learning architecture.

A. Event generation

We generate two data sets for the following studies, both using PYTHIA6. See Fig. 1 for an illustration of the two processes.

First, we generate jet samples using leading-order DIS as the hard-scattering process for the jet flavor tagging studies discussed in section IV. At leading order, the final state consists of the scattered electron and a single jet originating from different quark flavors. The leading-order DIS process is given the process number 99 according to PYTHIA6. We then identify the jet flavor with the flavor of the underlying quark in the leading-order DIS process ($q \to q$). We require the photon virtuality and inelasticity to be in the intervals $25 < Q^2 < 1000$ GeV^2 and $0.1 < y < 0.85$, respectively.

Since gluons do not contribute at leading order in DIS, we generate a second data set for quark vs. gluon jet tagging using di-jet events in low-Q^2 photoproduction events, including both the direct and resolved contributions. At leading order, the final state consists of the scattered electron in the forward direction close to the beam axis and a di-jet pair, which can be initiated by both quarks and gluons. We require low $10^{-5} < Q^2 < 1$ GeV^2, while maintaining the same cut on the inelasticity $0.1 < y < 0.85$. We identify quark and gluon jets in the photoproduction events using the PYTHIA6 resolved processes $11\,(qq \to qq), 12\,(q\bar{q} \to q\bar{q}), 53\,(gg \to q\bar{q})$ and $13\,(q\bar{q} \to gg), 68\,(gg \to gg)$ and the direct photon-gluon fusion processes $135\,(\gamma_T^* g \to q\bar{q}), 136\,(\gamma_L^* g \to q\bar{q})$. In order to avoid ambiguity in labeling processes with qg final states, we neglect the resolved process $28\,(qg \to qg)$ and the direct QCDC processes 131 $(\gamma_T^* q \to qg)$ and 132 $(\gamma_L^* q \to qg)$ and save these for future study.

In accordance with experimental particle detection capabilities, we include all particles in the event and in the jet reconstruction with a lifetime of $c\tau > 1$ cm. This includes

$$\gamma,\, e^-,\, \mu^-,\, \pi^-, p,\, n,\, K_L^0,\, K_S^0,\, K^-,\\ \Lambda^0,\, \Xi^0,\, \Xi^-,\, \Sigma^\pm,\, \Omega^-\,, \tag{10}$$

and the corresponding anti-particles. Particles with $c\tau < 1$ cm, such as neutral pions π^0, are decayed until daughters with $c\tau > 1$ cm are produced. The scattered electron is identified as the leading electron in the event and removed before we run the jet clustering algorithm. From the scattered electron we determine the virtuality Q^2 of the exchanged photon. We leave the implementation of a simulated detector response for future work, and as a first step we will instead examine the impact of particle identification information, charge information, and minimum transverse momentum requirements on the performance of these taggers in Sections IV and V.

Jets are reconstructed in the laboratory frame with the anti-k$_T$ algorithm [69]. We choose a jet radius parameter of $R = 1.0$ and we consider the rapidity range $|\eta_{\text{lab}}| < 4$. For the single jets identified in the DIS events, we require the transverse momentum of the identified jets to be $p_T > 10$ GeV. We consider a sample of $14M$ events satisfying these criteria, corresponding to an integrated luminosity of approximately 3 fb^{-1} (Check the number of jets per inverse femtobarn (see discussion w

Brian)), corresponding to approximately a few months of EIC runtime (estimated at approximately 10 fb^{-1} per year [1]. For the di-jets identified in the low-Q^2 photoproduction events, we require the transverse momentum of the leading jet to be $p_T > 8$ GeV and the subleading jet to be $p_T > 5$ GeV. We consider a sample of approximately $0.3M$ events satisfying these criteria, corresponding to an integrated luminosity of approximately XXXX fb^{-1} (Check the number of di-jets per inverse femtobarn). The size of the sample is determined by studying when the classification performance approximately saturates as the statistics are increased. We adopt a "UV definition" of the jet flavor content throughout this paper analogous to various machine learning studies of jet classification at the LHC.

We use supervised training throughout this work. In order to deploy supervised models on real data, one can either perform training on simulations or on experimental data itself. Training models on Monte Carlo simulations may be viable at the EIC if event simulations based on parton showers are sufficiently reliable in the sense that the machine learning algorithms can be trained on a suite of simulations and deployed on data. On the other hand, related data driven methods [70] and weakly supervised learning [71] may be suitable at the EIC as well. In addition, some of the studies proposed here can be performed on data before unfolding of detector effects. This will allow for the identification of suitable observable for which unfolding can be performed afterwards. Furthermore, recent advances in unfolding methods may allow one to unfold entire events [59]. We leave a more detailed exploration of these aspects for future work.

B. Machine learning algorithms

In the next sections, we will study the binary classification of jets with different quark flavor and quark vs. gluon jets using machine learning algorithms. For this task we choose deep sets [72–74] as our default classifier, which were introduced as Particle Flow Networks (PFNs) [15] for data obtained in high-energy collisions in particle and nuclear physics. The information about the particles in a jet or collider event can be considered as a set of four vectors with variable length event-by-event. A deep set / PFN is a neural network, which is invariant with respect to permutations of the input variables and it can naturally handle input with different length. This choice of machine learning architecture appears to be natural for data in collider physics since the number of particles varies event-by-event and there is no inherent ordering of the particles inside a jet or of the particles in the entire event. We note that other machine learning based classifiers were found to perform similarly or worse for analogous tasks at LHC energies [15, 16]. The PFNs take as input the information of all the particles inside a reconstructed jet. We represent the per-particle input variables as

$$p_i = (z_i, \eta_i, \phi_i, \text{PID}_i)\,, \tag{11}$$

where $z_i = p_{Ti}/p_T^{\text{jet}}$ is the normalized transverse momentum of particle i with respect to the beam axis, and η_i, ϕ_i are its rapidity and azimuthal angle. Following Ref. [15], we take η_i, ϕ_i relative to the (E-scheme [75]) jet axis. Lastly, PID_i in Eq. (11) denotes the particle identification number. See Eq. (10) for the different particles in our data set. We map the PIDs to numerical values in an interval around zero. The numerical values are separated by 0.1 and particles (anti particles) are assigned positive (negative) values. A deep set / PFN takes as input the kinematics of the N particles in the event $f(p_1, \ldots, p_N)$. It is constructed such that it satisfies $f(p_{\pi(1)}, \ldots, p_{\pi(N)})$, where π denotes the permutation operator. The required permutation invariance can be achieved by expressing f as [72]

$$f(p_1, \ldots, p_N) = F\left(\sum_{i=1}^{N} \Phi(p_i)\right). \tag{12}$$

Here Φ, F denote fully connected feed-forward neural networks with a certain number of hidden layers. The connections between hidden layers are parametrized in terms of weights and each node has a bias term. The per-particle neural network $\Phi : \mathbb{R}^4 \to \mathbb{R}^d$ maps the input to a d-dimensional latent space. The summation operation in latent space leads to the permutation invariance of f. The second neural network is a map between the latent space and the final output of the binary classifier $F : \mathbb{R}^d \to \mathbb{R}$.

We parametrize the functions Φ and F in Eqs. (**??**) and (**??**) in terms of DNNs, using the `EnergyFlow` package [15] with `Keras` [76]/`TensorFlow` [77]. For Φ we use two hidden layers with 100 nodes each and a latent space dimension of $d = 256$. For F we include three layers with 100 nodes each. For each dense layer we use the ReLU activation function [78] and we use the softmax activation function for the final output layer of the classifier. We train the neural networks using the Adam optimizer [79] and the binary cross entropy loss function [80], and train for 10 epochs with a batch size of 500. We reserve 20% of the training sample as a validation set, and an additional 20% as a test set on which all metrics are reported. We train the models using an NVIDIA A100 GPU on the Perlmutter supercomputer [].

The performance of a classifier is typically assessed by analyzing the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC). The ROC curve shows the cumulative distribution functions of the true positive rate vs. the false positive rate as the decision threshold is varied. In our case we define "positive" to refer to e.g. u jets. A random classifier follows a diagonal line with AUC$= 0.5$ and the better a classifier is, the closer the curve is to the upper left edge of the plot, with a perfect classifier having AUC$= 1$.

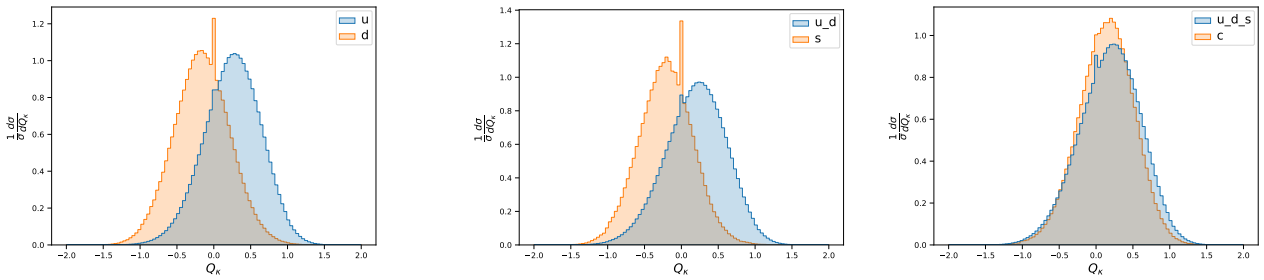

FIG. 2. The jet charge distribution for EIC jets with $p_T > 10$ GeV produced in high-Q^2 events as shown on the left side of Fig. 1. The three panels show the results for different flavor discrimination: u vs. d (left), ud vs. s (middle), and uds vs. c (right) for a jet charge parameter of $\kappa = 0.5$, see Eq. (13). The jet charge is able to distinguish u from d, s reasonably well, whereas it is a relatively poor discriminator for u vs. c or q vs. g. Note that a peak at $Q_\kappa = 0$ arises from jets that contain only neutral particles.

IV. JET FLAVOR TAGGING

Using the LO DIS events described in Sec. III A, we now study various binary classifications of quark-jet flavors. We consider several different classification groupings: u vs. d, ud vs. s, and uds vs. c. We will study the role of PID information, charge information, and minimum particle transverse momentum thresholds on the performance of the classifiers, as well as the role of both in-jet and out-of-jet particles.

We will benchmark our ML-based algorithms against the energy-weighted jet charge [81]

$$Q_\kappa = \sum_{i \in \text{jet}} z_i^\kappa Q_i \,, \tag{13}$$

where $z_i = p_{Ti}/p_T^{\text{jet}}$ denotes the longitudinal momentum fraction of the hadrons i inside the jet and Q_i is their electric charge. The weighting factor z_i^κ reduces the sensitivity to experimental uncertainties and κ is a free parameter that we will vary in our numerical studies below. The jet charge is soft safe but collinear unsafe, which means that theoretical calculations require a nonperturbative input that needs to be determined from experiment. Theoretical calculations of the jet charge were performed in Ref. [22, 23]. Extensions of the jet charge definition in Eq. (13) were proposed in Refs. [21, 82]. Experimental measurements at the LHC can be found in Refs. [83–85]. In Fig. 2, we show the jet charge distributions for the LO DIS jets considered in this section. The jet charge is able to distinguish u from d, s reasonably well, whereas it is a relatively poor discriminator for u vs. c since they have the same electric charge, and similarly for q vs. g (not shown here). The jet charge thereby serves as a reference to which the performance of our ML-based algorithms can be compared.

In order to study the role of PID information and charge information, we consider three variations of the information input to the PFN training:

- "PFN with PID": $p_i = (z_i, \eta_i, \phi_i, \text{PID}_i)$,
- "PFN with charge": $p_i = (z_i, \eta_i, \phi_i, Q_i)$,
- "PFN without PID": $p_i = (z_i, \eta_i, \phi_i)$.

We note that the "PFN with charge" classifier uses the same experimental information as the jet charge, whereas the "PFN with charge" uses full PID information, which is not used by the jet charge. Similarly, we consider varying the minimum transverse momentum of jet constituents input to the PFN training, varying between $p_T = 0.1 - 0.4$ GeV. While we do not consider the exact PID capabilities or single-particle efficiencies of the proposed EIC detectors, these variations provide a first-order estimate of the importance of PID and minimum particle transverse momentum detection capabilities and serve as an initial quantification of the value that may or may not be gained in jet tagging performance by investing in improved PID or minimum particle transverse momentum capabilities.

A. u vs. d

To begin, we consider the classification of u vs. d initiated jets. Our results are shown in Fig. 3. We find that while the jet charge is a fairly good discriminator of u vs. d jets, the PFN (which uses the full four-vector information of the final-state particles) improves the performance when either charge information is included or even more so when PID information is included. When neither PID nor charge information is included, the classifier cannot significantly distinguish u jets from d jets in PYTHIA6. Interestingly, the increase in performance when adding PID information rather than charge information is rather modest, considering the substantial experimental effort required to assign PID information to all constituents, especially noting that experimental PID capabilities are not perfectly efficient as assumed in our studies.

Next, we consider the role of the minimum transverse momentum of jet constituents input to the PFN training. Figure 4 shows the results when varying the minimum threshold between $p_{T,\text{particle}} > 0.1 - 0.4$ GeV. We find

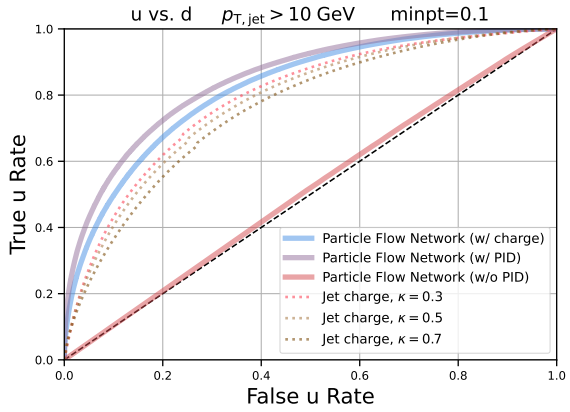

FIG. 3. ROC curve for u vs. d jet flavor tagging using the jet charge and PFNs for jets with $p_T > 10$ GeV and $p_{T,\text{particle}} > 0.1$ GeV. We consider three variations of the input to the PFN, providing either PID information for all particles, charge information for all particles, or neither.

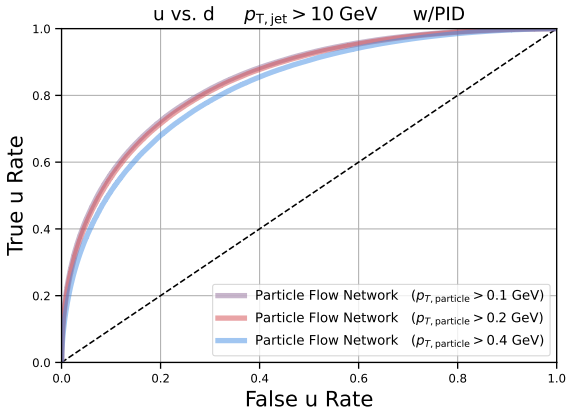

FIG. 4. ROC curves for u vs. d jet flavor tagging using using PFNs with PID information for jets with $p_T > 10$ GeV and different cuts on the minimum $p_{T,\text{particle}}$ required of jet constituents.

only a minor difference in the classifier performance when varying the minimum $p_{T,\text{particle}}$ between 0.1 GeV and 0.4 GeV, suggesting that the minimum $p_{T,\text{particle}}$ detector requirements are not essential for classifying jet flavor using the in-jet information. We will see, however, in the next section that this has a stronger impact when considering the out-of-jet particles.

B. Out-of-jet information

In this section, we investigate how the performance can be improved by not only making use of the particles inside the jet but also out-of-jet particles to classify the jet flavor. While we have used a relatively large jet radius $R = 1.0$ in these studies, this choice is somewhat arbitrary and neglects the role of large-angle radiation and correlations across the entire event. We therefore compare the performance of a PFN supplied with only in-jet particles to that of a PFN supplied with both in-jet and out-of-jet particles. Figure 5 shows the results of this comparison. We show the comparison for two different minimum $p_{T,\text{particle}}$ thresholds, 0.1 GeV and 0.4 GeV We find that the difference between the in-jet classifier and the in-jet + out-of-jet classifier is significant for the case $p_{T,\text{particle}} > 0.1$ GeV, whereas the difference is almost negligible for $p_{T,\text{particle}} > 0.4$ GeV. This suggests that the *soft* out-of-jet particles play a significant role in boosting the classification performance – despite that the soft in-jet particles had little impact (see Fig. 4). This motivates further study of the origin and role of out-of-jet radiation, since our results suggest it can provide a significant boost in jet flavor tagging performance.

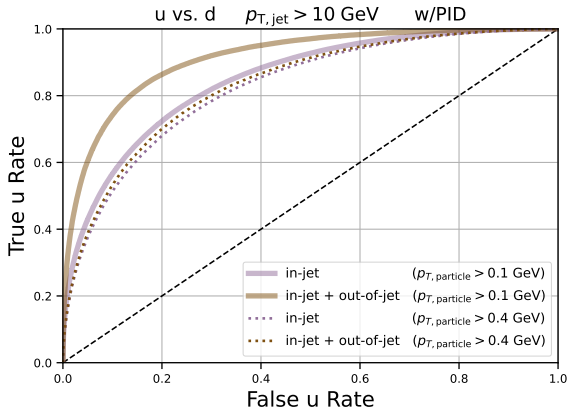

FIG. 5. ROC curves for u vs. d jet flavor tagging using PFNs with PID information for jets with $p_T > 10$ GeV, using either in-jet information as input or using both in-jet and out-of-jet information as input. We consider two different cuts on the minimum p_T required of jet constituents, which illustrate that soft out-of-jet particles play a significant role in boosting the classification performance.

C. Strange and charm

We now turn to the identification of strange- and charm-quark initiated jets. Since strange- and charm-initiated jets are considerably more rare than up- or down-initiated jets (for our kinematics, the relative u:d:s:c ratios are approximately 33:5:2:1), we quantify the classification performance using both the ROC curve and the precision-recall curve. In fact, strange jets are even more rare than charm jets, since despite that the proton PDF contains a larger quantity of strange than charm, the overall cross section for charm is larger due to its larger electric charge.

Strange and charm jets also differ from up and down jets in that strange and charm hadrons have limited decay

lifetimes. In the case of strange quarks, there are a variety of weakly decaying strange hadrons with lifetimes 1 cm $< c\tau <$ 10 cm (namely K^0_S, Λ^0, Ξ^0, Ξ^-, $\Sigma^\pm$, Ω^- and their associated antiparticles) which therefore decay on a length scale comparable to the size of the innermost tracking layers of collider experiments. We therefore will contrast the classification performance depending on whether the PFN is provided the undecayed strange hadrons or only the decay products of these hadrons. In the case of charm quarks, on the other hand, all charm hadrons decay with lifetimes much shorter than $c\tau = 1$ cm, and cannot be directly detected by experiments but rather must be reconstructed using the invariant mass of decay products of exclusive charm hadrons or by tagging displaced vertices. A large literature exists on charm-jet tagging algorithms, which we do not pursue further here [].

Figure 6 shows the results for ud vs. s jet classification with final-state particle decay lifetimes of $c\tau >$1cm and $c\tau >$ 10cm, respectively. We find several notable differences compared to the u vs. d classification. First, the PFN with PID dramatically outperforms the jet charge. We also provide as a reference a simple "strange tagger" which classifies the jet flavor purely based on whether a strange hadron is included in the jet. The PFN also dramatically outperforms this. This provides a clear illustration of the value of ML-based jet flavor identification. Second, the overall performance of ud vs. s tagging is significantly improved when PID information is provided relative to charge information, especially when the weakly decaying strange hadrons with $c\tau > 1$ cm are included as input to the PFN. If only charge information is supplied, the performance decreases substantially. This provides a clear illustration that PID information is highly valuable to obtaining the best possible strange-jet tagging performance. We leave further study, such as whether providing PID information of the leading particle rather than all particles, which could substantially lessen the experimental efforts, to future work. Third, if neither PID nor charge information is provided, the performance is yet again substantially worse – however it is still notably better than in the u vs. d case. This illustrate the importance of particle identification vs. fragmentation in determining the jet flavor – since when neither PID nor charge information is provided the ML algorithm can only learn from the differences in fragmentation between ud and s jets.

Figure 7 shows the results for uds vs. c jet classification. In this case, the jet charge is not expected to be a good discriminator, since u (which dominates the uds sample) and c jets have the same electric charge. We find similarly strong performance of the PFN classifier when PID information is included, with an even larger benefit of providing PID information relative to charge information. Additionally, we note that the PFN that is supplied with neither PID nor charge information performs better than the previous cases, illustrating that the amount of information in the fragmentation pattern unrelated to particle PID or charge plays an increasing role for heavier quarks.

V. HARD-SCATTERING EVENT TAGGING

The motivation of machine learned-jet classification at the EIC and RHIC is quite different compared to the LHC. For example, at the LHC dijet reference processes can be used as calibration and the resulting classifier can be applied to identify jets in multi-jet events to search for physics beyond the Standard Model. Instead, at RHIC and the EIC the focus will be on improving for example measurements of spin asymmetries as discussed below or to improve constraints on cold nuclear matter effects. Therefore, at RHIC and the EIC, the classifier does not need to be limited to the particles inside the identified jet. Often, the classification of the hard-scattering event is of primary interest instead of the classification of a single jet. Therefore, we propose to not only utilize the particles inside the reconstructed jet but to also take as input particles outside the jet. Note that we still require a jet with a given transverse momentum to identify the entire event to ensure the presence of a hard-scale, which allows for the interpretation or applicability of perturbative techniques in QCD. We illustrate the difference of jet vs. event classification for the leading-order DIS process in Fig. 8. We expect that the additional information contained in the dynamics of particles outside the reconstructed jet can generally increase the performance of the machine learning algorithm. We note that event type classification using triggers and machine learning was discussed in Ref. [4] and references therein. Different than Ref. [4], we aim here at identifying the hard-scattering event at parton level. As discussed in section III B above, the in-jet information that is used to train machine learned classifiers can be captured by complete sets of observables like N-subjettiness and EFPs. Similar observable bases can be constructed for out-of-jet information and correlations between jets (such as in photoproduction events, see Fig. 1 can be captured by observables like the jet pull [25, 26].

We consider two examples of event classification in this section. In both cases, we use low-Q^2 photoproduction events that contain a di-jet with the transverse momentum of the leading jet required to be $p_T > 8$ GeV and the subleading jet to be $p_T > 5$ GeV, as described in Sec. III A. First, we consider the classification quark vs. gluon jet topologies by discriminating qq or $q\bar{q}$ di-jet topologies from gg topologies. Second, we consider the classification of direct vs. resolved photoproduction processes. Similar to the in-jet particles, we normalize the transverse momentum of out-of-jet particles relative to the leading jet transverse momentum $z_i = p_{Ti}/p^{\text{jet}}_{T1}$. Since we divide by the transverse momentum of the leading jet in the event, we have $z_i < 1$ for both in-jet and out-of-jet particles. Moreover, we count the values (η_i, ϕ_i) of out-of-jet particles relative to the leading jet axis.

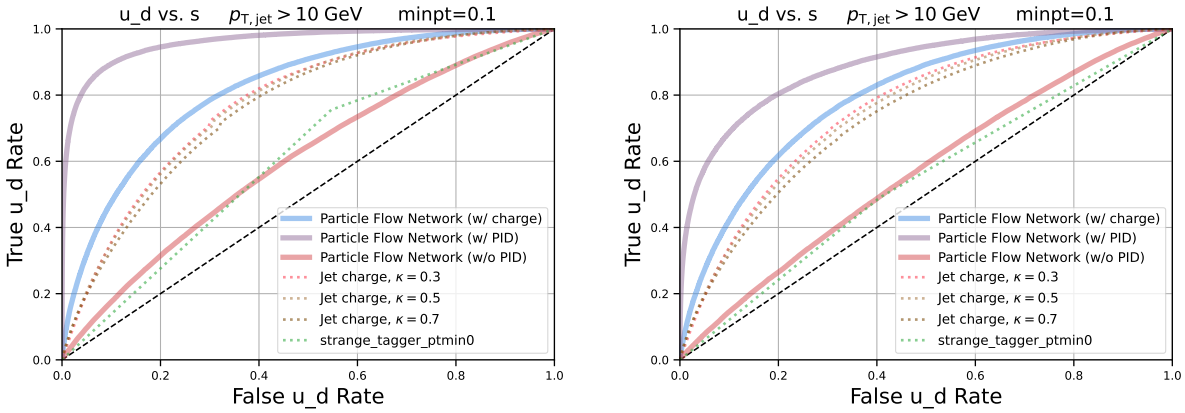

FIG. 6. ROC curves for ud vs s jet flavor tagging using the jet charge and PFNs for jets with $p_T > 10$ GeV and $p_{T,\text{particle}} > 0.1$ GeV. The left panel is constructed from particles with a decay length $c\tau > 1$ cm (in which the weakly-decaying strange hadrons K^0_S, Λ^0, Ξ^0, Ξ^-, $\Sigma^\pm$, Ω^- and their associated antiparticles are undecayed), whereas the right panel is constructed from particles with a decay length $c\tau > 10$ cm (in which the above weakly-decaying strange hadrons are decayed). We consider three variations of the input to the PFN, providing either PID information for all particles, charge information for all particles, or neither.

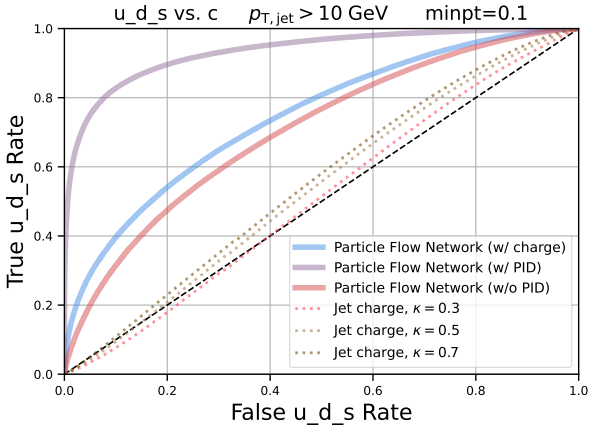

FIG. 7. ROC curves for uds vs. c jet flavor tagging using the jet charge and PFNs for jets with $p_T > 10$ GeV and $p_{T,\text{particle}} > 0.1$ GeV. We consider three variations of the input to the PFN, providing either PID information for all particles, charge information for all particles, or neither.

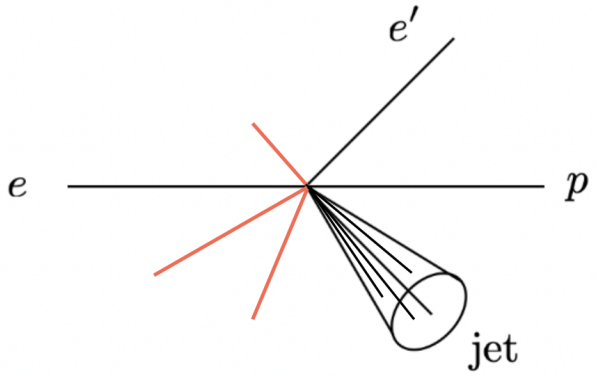

FIG. 8. Illustration of particles inside the jet (black) and out-of-jet radiation (red), which we also take into account to classify the hard-scattering event as discussed in section V. As an example, we show jet production in a high-Q^2 DIS scattering process.

A. Quark vs. gluon jet tagging

We consider events with quark and gluon di-jet topologies by considering both direct and resolved processes that result in qq-, $q\bar{q}$-, or gg-initiated di-jets, as described in Sec. III A. We then train PFNs using either (i) the leading jet particles, (ii) both the leading and subleading jet particles, or (iii) All particles in jets with $p_{T,\text{jet}} > 2$ GeV.

Figure 9 shows the classification performance of quark vs. gluon jet event topologies. The PFN trained with only the leading jet particles corresponds most closely to previous studies of the classification of quark vs. gluon single jets[]. While the performance at the low EIC jet energies considered here is lower than quark vs. gluon classification with high-p_T jets at the LHC, the PFN still is able to achieve substantial classification performance. As the subleading jet particles and out-of-di-jet particles are added to the PFN training input, the performance gradually increases.

B. Direct vs. resolved processes and improved constraints on photon structure

Next, we consider discriminating events that arise from direct vs. resolved photoproduction processes. We con-

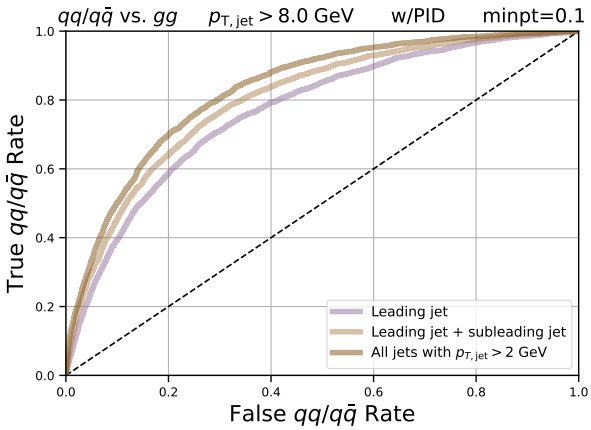

FIG. 9. ROC curves for quark vs. gluon event tagging at the EIC with PFNs including PID information and $p_{T,\text{particle}} > 0.1$ GeV. Here we consider quark and gluon jets produced in low-Q^2 protoproduction events containing di-jets with leading jet required to be $p_T > 8$ GeV and the subleading jet to be $p_T > 5$ GeV, see Fig. 1.

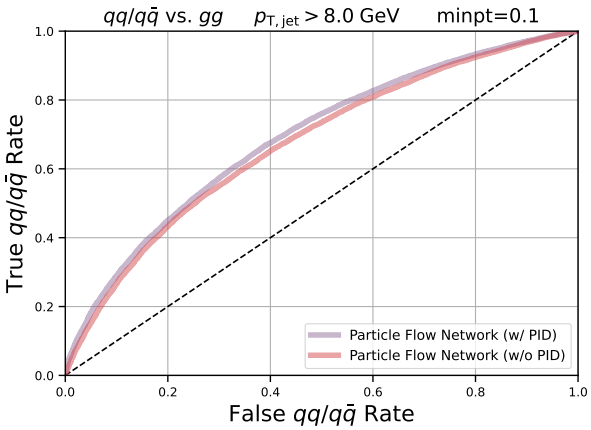

FIG. 10. ROC curves for direct vs. resolved process tagging at the EIC. Here we consider quark and gluon jets produced in low-Q^2 photoproduction events, see Fig. 1.

sider events with the same quark and gluon di-jet topologies described in the previous section except additionally including qg-initiated di-jets in addition to qg-, $q\bar{q}$-, and gg-initiated di-jets. The direct processes correspond to those initiated by the virtual photon, whereas the resolved processes are initiated by partons emanating from the virtual photon. In this way, classifying direct vs. resolved processes can improve constraints on the photon structure, in particular the parton-in-photon PDF. (Mention also polarized parton-in-photon PDF, see Elke's paper [86**? ? ? ?**].)

Figure 10 shows the classification performance of direct vs. resolved photoproduction processes. We find that the performance is worse than the quark vs. gluon di-jet topology classification shown in the previous section, which is unsurprising given that the direct and resolved contributions both contain both quark and gluon jets. We furthermore find that the impact of supplying PID information to the PFN is almost negligible in this case.

VI. CONCLUSIONS AND OUTLOOK

We have presented first studies of machine learning based jet and event classification using simulated events at the Electron-Ion Collider. While the performance of jet flavor classification is more challenging than with high transverse momentum jets at the LHC, ML-based classification algorithms offer important advantages in performance and many prospects for interpretability. We found that machine learning algorithms outperform traditional observables used to identify jet flavor, such as the jet charge. In order to provide input to the detector specifications at the EIC, we investigated the impact of PID information, charge information, and minimum particle transverse momentum requirements. We found that providing charge information is sufficient for u vs. d jet classification, but that PID information gives large improvement to strange and charm jet tagging capability. We found that soft particles with $0.1 < p_T < 0.4$ GeV have only minor impact when jet flavor classification is performed using in-jet particles, but that out-of-jet soft particles give substantial improvement to the classification performance.

These methods will play an important role in multiple areas of the EIC science program. Jet flavor tagging can lead to improved constraints of transverse momentum dependent parton distribution functions, and hard event tagging provides opportunities to increase experimental access to transverse single spin asymmetries. Additionally, the methods outlined here can be applied to classify ep vs. eA collisions and provide new insight to cold nuclear matter effects, and can be extended to produce increasingly interpretable results []. The studies performed here can be extended to a full detector simulation as well as additional Monte Carlo event generators and can eventually potentially be deployed on experimental data itself.

DATA AVAILABILITY

ACKNOWLEDGMENTS

We would like to thank Brian Page, Nobuo Sato ...? for helpful discussions. FR was supported by the Simons Foundation under the Simons Bridge program for Postdoctoral Fellowships at SCGP and YITP award number 815892; the NSF, award number 1915093; the DOE Contract No. DE-AC05-06OR23177, under which Jefferson Science Associates, LLC operates Jefferson Lab and Old Dominion University. This research used resources of the

National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

[1] R. Abdul Khalek *et al.*, Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report, (2021), arXiv:2103.05419 [physics.ins-det].

[2] M. Arratia, Y. Song, F. Ringer, and B. V. Jacak, Jets as precision probes in electron-nucleus collisions at the future Electron-Ion Collider, Phys. Rev. C **101**, 065204 (2020), arXiv:1912.05931 [nucl-ex].

[3] B. S. Page, X. Chu, and E. C. Aschenauer, Experimental Aspects of Jet Physics at a Future EIC, Phys. Rev. D **101**, 072003 (2020), arXiv:1911.00657 [hep-ph].

[4] A. Boehnlein *et al.*, Artificial Intelligence and Machine Learning in Nuclear Physics, (2021), arXiv:2112.02309 [nucl-th].

[5] M. Feickert and B. Nachman, A Living Review of Machine Learning for Particle Physics, (2021), arXiv:2102.02770 [hep-ph].

[6] Y. Alanazi *et al.*, Simulation of electron-proton scattering events by a Feature-Augmented and Transformed Generative Adversarial Network (FAT-GAN) 10.24963/ijcai.2021/293 (2020), arXiv:2001.11103 [hep-ph].

[7] Y. S. Lai, D. Neill, M. Płoskoń, and F. Ringer, Explainable machine learning of the underlying physics of high-energy particle collisions, (2020), arXiv:2012.06582 [hep-ph].

[8] A. Butter, T. Heimel, S. Hummerich, T. Krebs, T. Plehn, A. Rousselot, and S. Vent, Generative Networks for Precision Enthusiasts, (2021), arXiv:2110.13632 [hep-ph].

[9] M. Arratia, D. Britzger, O. Long, and B. Nachman, Reconstructing the kinematics of deep inelastic scattering with deep learning, Nucl. Instrum. Meth. A **1025**, 166164 (2022), arXiv:2110.05505 [hep-ex].

[10] M. Diefenthaler, A. Farhat, A. Verbytskyi, and Y. Xu, Deeply Learning Deep Inelastic Scattering Kinematics, (2021), arXiv:2108.11638 [hep-ph].

[11] R. D. Ball *et al.* (NNPDF), Parton distributions for the LHC Run II, JHEP **04**, 040, arXiv:1410.8849 [hep-ph].

[12] P. Gras, S. Höche, D. Kar, A. Larkoski, L. Lönnblad, S. Plätzer, A. Siódmok, P. Skands, G. Soyez, and J. Thaler, Systematics of quark/gluon tagging, JHEP **07**, 091, arXiv:1704.03878 [hep-ph].

[13] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, Jet-images — deep learning edition, JHEP **07**, 069, arXiv:1511.05190 [hep-ph].

[14] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, Deep learning in color: towards automated quark/gluon jet discrimination, JHEP **01**, 110, arXiv:1612.01551 [hep-ph].

[15] P. T. Komiske, E. M. Metodiev, and J. Thaler, Energy Flow Networks: Deep Sets for Particle Jets, JHEP **01**, 121, arXiv:1810.05165 [hep-ph].

[16] Y. Lu, A. Romero, M. J. Fenton, D. Whiteson, and P. Baldi, Resolving extreme jet substructure, JHEP **08**, 046, arXiv:2202.00723 [hep-ex].

[17] P. T. Komiske, E. M. Metodiev, and J. Thaler, Energy flow polynomials: A complete linear basis for jet substructure, JHEP **04**, 013, arXiv:1712.07124 [hep-ph].

[18] K. Datta and A. Larkoski, How Much Information is in a Jet?, JHEP **06**, 073, arXiv:1704.08249 [hep-ph].

[19] K. Datta and A. J. Larkoski, Novel Jet Observables from Machine Learning, JHEP **03**, 086, arXiv:1710.01305 [hep-ph].

[20] K. Datta, A. Larkoski, and B. Nachman, Automating the Construction of Jet Observables with Machine Learning, Phys. Rev. D **100**, 095016 (2019), arXiv:1902.07180 [hep-ph].

[21] K. Fraser and M. D. Schwartz, Jet Charge and Machine Learning, JHEP **10**, 093, arXiv:1803.08066 [hep-ph].

[22] W. J. Waalewijn, Calculating the Charge of a Jet, Phys. Rev. D **86**, 094030 (2012), arXiv:1209.3019 [hep-ph].

[23] D. Krohn, M. D. Schwartz, T. Lin, and W. J. Waalewijn, Jet Charge at the LHC, Phys. Rev. Lett. **110**, 212001 (2013), arXiv:1209.2421 [hep-ph].

[24] I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, N-Jettiness: An Inclusive Event Shape to Veto Jets, Phys. Rev. Lett. **105**, 092002 (2010), arXiv:1004.2489 [hep-ph].

[25] J. Gallicchio and M. D. Schwartz, Seeing in Color: Jet Superstructure, Phys. Rev. Lett. **105**, 022001 (2010), arXiv:1001.5027 [hep-ph].

[26] A. J. Larkoski, S. Marzani, and C. Wu, Theory Predictions for the Pull Angle, Phys. Rev. D **99**, 091502 (2019), arXiv:1903.02275 [hep-ph].

[27] M. Arratia, Y. Furletova, T. J. Hobbs, F. Olness, and S. J. Sekula, Charm jets as a probe for strangeness at the future Electron-Ion Collider, Phys. Rev. D **103**, 074023 (2021), arXiv:2006.12520 [hep-ph].

[28] S. Caletti, O. Fedkevych, S. Marzani, and D. Reichelt, Tagging the initial-state gluon, Eur. Phys. J. C **81**, 844 (2021), arXiv:2108.10024 [hep-ph].

[29] A. Banfi, G. P. Salam, and G. Zanderighi, Infrared safe definition of jet flavor, Eur. Phys. J. C **47**, 113 (2006), arXiv:hep-ph/0601139.

[30] A. Banfi, G. P. Salam, and G. Zanderighi, Accurate QCD predictions for heavy-quark jets at the Tevatron and LHC, JHEP **07**, 026, arXiv:0704.2999 [hep-ph].

[31] S. Caletti, A. J. Larkoski, S. Marzani, and D. Reichelt, Practical jet flavour through NNLO, Eur. Phys. J. C **82**, 632 (2022), arXiv:2205.01109 [hep-ph].

[32] S. Caletti, A. J. Larkoski, S. Marzani, and D. Reichelt, A Fragmentation Approach to Jet Flavor, (2022), arXiv:2205.01117 [hep-ph].

[33] R. Gauld, A. Huss, and G. Stagnitto, A dress of flavour to suit any jet, (2022), arXiv:2208.11138 [hep-ph].

[34] Y. S. Lai, J. Mulligan, M. Płoskoń, and F. Ringer, The information content of jet quenching and machine learning assisted observable design, (2021), arXiv:2111.14589 [hep-ph].

[35] M. Cranmer, A. Sanchez-Gonzalez, P. Battaglia, R. Xu, K. Cranmer, D. Spergel, and S. Ho, Discovering symbolic models from deep learning with inductive biases, NeurIPS 2020 (2020), arXiv:2006.11287 [cs.LG].

[36] J. Adam *et al.* (STAR), Measurement of transverse single-spin asymmetries of π^0 and electromagnetic jets at forward rapidity in 200 and 500 GeV transversely polarized proton-proton collisions, Phys. Rev. D **103**, 092009 (2021), arXiv:2012.11428 [hep-ex].

[37] B. I. Abelev *et al.* (STAR), Measurement of transverse single-spin asymmetries for di-jet production in proton-proton collisions at s**(1/2) = 200-GeV, Phys. Rev. Lett. **99**, 142003 (2007), arXiv:0705.4629 [hep-ex].

[38] Azimuthal transverse single-spin asymmetries of inclusive jets and identified hadrons within jets from polarized pp collisions at $\sqrt{s}$ = 200 GeV, (2022), arXiv:2205.11800 [hep-ex].

[39] U. A. Acharya *et al.* (PHENIX), Transverse-single-spin asymmetries of charged pions at midrapidity in transversely polarized p+p collisions at $\sqrt{s} = 200$ GeV, Phys. Rev. D **105**, 032003 (2022), arXiv:2112.05680 [hep-ex].

[40] U. A. Acharya *et al.* (PHENIX), Probing Gluon Spin-Momentum Correlations in Transversely Polarized Protons through Midrapidity Isolated Direct Photons in $p^{\uparrow}$+p Collisions at $\sqrt{s}$=200 GeV, Phys. Rev. Lett. **127**, 162001 (2021), arXiv:2102.13585 [hep-ex].

[41] U. A. Acharya *et al.* (PHENIX), Improving constraints on gluon spin-momentum correlations in transversely polarized protons via midrapidity open-heavy-flavor electrons in $p^{\uparrow} + p$ collisions at $\sqrt{s} = 200$ GeV, (2022), arXiv:2204.12899 [hep-ex].

[42] D. de Florian, R. Sassot, M. Stratmann, and W. Vogelsang, Evidence for polarization of gluons in the proton, Phys. Rev. Lett. **113**, 012001 (2014), arXiv:1404.4293 [hep-ph].

[43] E. R. Nocera, R. D. Ball, S. Forte, G. Ridolfi, and J. Rojo (NNPDF), A first unbiased global determination of polarized PDFs and their uncertainties, Nucl. Phys. B **887**, 276 (2014), arXiv:1406.5539 [hep-ph].

[44] Y. Zhou, N. Sato, and W. Melnitchouk (Jefferson Lab Angular Momentum (JAM)), How well do we know the gluon polarization in the proton?, Phys. Rev. D **105**, 074022 (2022), arXiv:2201.02075 [hep-ph].

[45] U. A. Acharya *et al.* (PHENIX), Measurement of charged pion double spin asymmetries at midrapidity in longitudinally polarized $p + p$ collisions at $\sqrt{s} = 510$ GeV, Phys. Rev. D **102**, 032001 (2020), arXiv:2004.02681 [hep-ex].

[46] M. S. Abdallah *et al.* (STAR, (STAR Collaboration)†), Longitudinal double-spin asymmetry for inclusive jet and dijet production in polarized proton collisions at $\sqrt{s} = 510$ GeV, Phys. Rev. D **105**, 092011 (2022), arXiv:2110.11020 [hep-ex].

[47] L. Zheng, E. C. Aschenauer, J. H. Lee, B.-W. Xiao, and Z.-B. Yin, Accessing the gluon Sivers function at a future electron-ion collider, Phys. Rev. D **98**, 034011 (2018), arXiv:1805.05290 [hep-ph].

[48] V. Cirigliano, K. Fuyuto, C. Lee, E. Mereghetti, and B. Yan, Charged Lepton Flavor Violation at the EIC, JHEP **03**, 256, arXiv:2102.06176 [hep-ph].

[49] R. Boughezal, A. Emmert, T. Kutz, S. Mantry, M. Nycz, F. Petriello, K. Şimşek, D. Wiegand, and X. Zheng, Neutral-current electroweak physics and SMEFT studies at the EIC, Phys. Rev. D **106**, 016006 (2022), arXiv:2204.07557 [hep-ph].

[50] J. L. Zhang *et al.*, Search for $e \to \tau$ Charged Lepton Flavor Violation at the EIC with the ECCE Detector, (2022), arXiv:2207.10261 [hep-ph].

[51] H. T. Li, B. Yan, and C. P. Yuan, Jet Charge: A new tool to probe the anomalous $Zb\bar{b}$ couplings at the EIC, (2021), arXiv:2112.07747 [hep-ph].

[52] Y. Hatta, N. Mueller, T. Ueda, and F. Yuan, QCD Resummation in Hard Diffractive Dijet Production at the Electron-Ion Collider, Phys. Lett. B **802**, 135211 (2020), arXiv:1907.09491 [hep-ph].

[53] Y. Hatta, B.-W. Xiao, F. Yuan, and J. Zhou, Anisotropy in Dijet Production in Exclusive and Inclusive Processes, Phys. Rev. Lett. **126**, 142001 (2021), arXiv:2010.10774 [hep-ph].

[54] Y. Hatta, B.-W. Xiao, F. Yuan, and J. Zhou, Azimuthal angular asymmetry of soft gluon radiation in jet production, Phys. Rev. D **104**, 054037 (2021), arXiv:2106.05307 [hep-ph].

[55] A. Kerbizi and L. Lönnblad, StringSpinner - adding spin to the PYTHIA string fragmentation, Comput. Phys. Commun. **272**, 108234 (2022), arXiv:2105.09730 [hep-ph].

[56] D. Boer and W. Vogelsang, Asymmetric jet correlations in p p uparrow scattering, Phys. Rev. D **69**, 094025 (2004), arXiv:hep-ph/0312320.

[57] X. Liu, F. Ringer, W. Vogelsang, and F. Yuan, Factorization and its Breaking in Dijet Single Transverse Spin Asymmetries in pp Collisions, Phys. Rev. D **102**, 114012 (2020), arXiv:2008.03666 [hep-ph].

[58] Z.-B. Kang, K. Lee, D. Y. Shao, and J. Terry, The Sivers Asymmetry in Hadronic Dijet Production, JHEP **02**, 066, arXiv:2008.05470 [hep-ph].

[59] A. Andreassen, P. T. Komiske, E. M. Metodiev, B. Nachman, and J. Thaler, OmniFold: A Method to Simultaneously Unfold All Observables, Phys. Rev. Lett. **124**, 182001 (2020), arXiv:1911.09107 [hep-ph].

[60] D. W. Sivers, Single Spin Production Asymmetries from the Hard Scattering of Point-Like Constituents, Phys. Rev. D **41**, 83 (1990).

[61] J. C. Collins, Fragmentation of transversely polarized quarks probed in transverse momentum distributions, Nucl. Phys. B **396**, 161 (1993), arXiv:hep-ph/9208213.

[62] A. Schafer and O. V. Teryaev, Sum rules for the T - odd fragmentation functions, Phys. Rev. D **61**, 077903 (2000), arXiv:hep-ph/9908412.

[63] S. Meissner, A. Metz, and D. Pitonyak, Momentum sum rules for fragmentation functions, Phys. Lett. B **690**, 296 (2010), arXiv:1002.4393 [hep-ph].

[64] K. Goeke, S. Meissner, A. Metz, and M. Schlegel, Checking the Burkardt sum rule for the Sivers function by model calculations, Phys. Lett. B **637**, 241 (2006), arXiv:hep-ph/0601133.

[65] M. Burkardt, Sivers mechanism for gluons, Phys. Rev. D **69**, 091501 (2004), arXiv:hep-ph/0402014.

[66] M. Burkardt, Quark correlations and single spin asymmetries, Phys. Rev. D **69**, 057501 (2004), arXiv:hep-ph/0311013.

[67] X. Liu and H. Xing, The Time-reversal Odd Side of a Jet, (2021), arXiv:2104.03328 [hep-ph].

[68] W. K. Lai, X. Liu, M. Wang, and H. Xing, Unveiling Nucleon 3D Chiral-Odd Structure with Jet Axes, (2022), arXiv:2205.04570 [hep-ph].

[69] M. Cacciari, G. P. Salam, and G. Soyez, The anti-k_t jet clustering algorithm, JHEP **04**, 063, arXiv:0802.1189 [hep-ph].

[70] E. M. Metodiev and J. Thaler, Jet Topics: Disentangling Quarks and Gluons at Colliders, Phys. Rev. Lett. **120**, 241602 (2018), arXiv:1802.00008 [hep-ph].

[71] L. M. Dery, B. Nachman, F. Rubbo, and A. Schwartzman, Weakly Supervised Classification in High Energy Physics, JHEP **05**, 145, arXiv:1702.00414 [hep-ph].

[72] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Póczos, R. Salakhutdinov, and A. J. Smola, Deep sets, CoRR **abs/1703.06114** (2017), arXiv:1703.06114.

[73] E. Wagstaff, F. B. Fuchs, M. Engelcke, I. Posner, and M. A. Osborne, On the limitations of representing functions on sets, CoRR **abs/1901.09006** (2019), arXiv:1901.09006.

[74] B. Bloem-Reddy and Y. Teh, Probabilistic symmetries and invariant neural networks, Journal of Machine Learning Research **21**, 1 (2020).

[75] G. C. Blazey *et al.*, Run II jet physics, in *Physics at Run II: QCD and Weak Boson Physics Workshop: Final General Meeting* (2000) pp. 47–77, arXiv:hep-ex/0005012.

[76] F. Chollet *et al.*, Keras, `https://github.com/fchollet/keras` (2015).

[77] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, and et al., TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from tensorflow.org.

[78] V. Nair and G. E. Hinton, Rectified linear units improve restricted boltzmann machines, in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, edited by J. Fürnkranz and T. Joachims (2010) pp. 807–814.

[79] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, CoRR **abs/1412.6980** (2015), arXiv:1412.6980.

[80] D. R. Cox, The regression analysis of binary sequences, Journal of the Royal Statistical Society: Series B (Methodological) **20**, 215 (1958).

[81] R. D. Field and R. P. Feynman, A Parametrization of the Properties of Quark Jets, Nucl. Phys. B **136**, 1 (1978).

[82] Z.-B. Kang, X. Liu, S. Mantry, M. C. Spraker, and T. Wilson, Dynamic Jet Charge, Phys. Rev. D **103**, 074028 (2021), arXiv:2101.04304 [hep-ph].

[83] G. Aad *et al.* (ATLAS), Measurement of jet charge in dijet events from $\sqrt{s}$=8 TeV pp collisions with the ATLAS detector, Phys. Rev. D **93**, 052003 (2016), arXiv:1509.05190 [hep-ex].

[84] A. M. Sirunyan *et al.* (CMS), Measurements of jet charge with dijet events in pp collisions at $\sqrt{s} = 8$ TeV, JHEP **10**, 131, arXiv:1706.05868 [hep-ex].

[85] D. A. Hangal (CMS), Studies of Quark and Gluon Contributions to Jets using Jet Charge Measurements in pp and PbPb Collisions, Nucl. Phys. A **1005**, 121803 (2021), arXiv:2004.14600 [nucl-ex].

[86] X. Chu, E.-C. Aschenauer, J.-H. Lee, and L. Zheng, Photon structure studied at an Electron Ion Collider, Phys. Rev. D **96**, 074035 (2017), arXiv:1705.08831 [nucl-ex].